1 # 6   ASSESS THE RESULTS OF THE DISPOSITION SURVEY

2 ## 6.1   Introduction

3 The assessment phase of the Data Life Cycle involves the interpretation of survey results.

4 Interpretation of survey results is very straightforward when all of the data are below or all of the

5 data are above the action level, and the correct decision regarding disposition of the M&E is

6 obvious. In these cases very little data interpretation is required. However, formal statistical

7 tests provide a valuable tool when the survey results are neither clearly above nor entirely below

8 the action level. In either case, statistical tests are always used to support the survey design in

9 helping to ensure the quantity and quality of data meet the data quality objectives (DQOs) and

10 measurement quality objectives (MQOs).

11 ## 6.2   Conduct Data Quality Assessment

12 Data Quality Assessment (DQA) is a scientific and statistical evaluation that determines whether

13 data are the right type, quality and quantity to support their intended use (EPA 2006b). There are

14 five steps in the DQA Process:

15   1. Review the DQOs and survey design.

16   2. Conduct a preliminary data review.

17   3. Select the statistical test.

18   4. Verify the assumptions of the statistical test.

19   5. Draw conclusions from the data.

20 The effort applied to DQA should be consistent with the graded approach used to develop the

21 survey design. More information on DQA can be found in *Data Quality Assessment: A User's*

22 *Guide* (EPA QA/G-9R, EPA 2006b) and *Data Quality Assessment: Statistical Tools for*

23 *Practitioner*s (EPA QA/G-9S, EPA 2006c). Data should be verified and validated as described

24 in the quality assurance project plan (QAPP). Guidance on data verification and validation can

25 be found in MARSSIM Section 9.3 and MARLAP Chapter 8. Guidance on developing a QAPP

26 is available in EPA QA/G-5 (EPA 2002a) and MARLAP Chapter 4.

27 **6.2.1   Review the Data Quality Objectives and Survey Design**

28   The first step in the DQA Process is a review of the DQO outputs used to develop the survey

29   design to ensure they are still applicable. The review of the DQOs and survey design should also

30   include the MQOs (e.g., measurement uncertainty, detectability, and quantifiability). For

31   example, if the data show the measurement uncertainty exceeds the estimate used to design the

32   survey, the DQOs and MQOs should be revisited.

33   The survey design should be reviewed for consistency with the DQOs. For example, the review

34   should verify that the appropriate number or amount of measurements were performed in the

35   correct locations and were analyzed using measurement methods with adequate sensitivity.

36   In cases where the survey did not involve taking discrete measurements or samples (i.e., scan-

37   only, conveyor systems, or in situ surveys), it is imperative that the minimum detectable

38   concentrations (MDCs) be calculated realistically and they truly reflect at least 95 percent

39   probability that concentrations at or about the MDC were detected. Clearly, MDCs must be

40   capable of detecting radionuclide concentrations or levels of radioactivity at or below the upper

41   bound of the gray region (UBGR). When detection decisions are made for individual items (i.e.,

42   Scenario B) the MDC should be less than or equal to the UBGR.

43   The minimum quantifiable concentration (MQC) is defined as the radionuclide concentration or

44   level of radioactivity at which the measurement method gives results with a specified relative

45   standard deviation $1/k_Q$, where k_Q is usually chosen to be 10 (see Section 5.8, MARLAP Section

46   19.4.5, MARLAP Section 19.7.3). MARSAME recommends that the MQC should be no larger

47   than the upper bound of the gray region (UBGR) when making quantitative comparisons of the

48   mean survey data to the action level (i.e., Scenario A). This is an expression of the fact that the

49   MQC, unlike the MDC used for a simple detection decision, addresses the relative uncertainty of

50   the data value obtained. If the objective of the disposition survey is to quantify radionuclide

51   concentrations near the UBGR, the MQC should be no larger than the UBGR.[1]

---

[1] The UBGR is either the action level for Scenario A or the discrimination limit for Scenario B (see Section 4.2).

52   For MARSSIM-type surveys (see Section 4.4.3) it is important to collect sufficient data to

53   support a disposition decision. This is particularly important in cases where the radionuclide

54   concentrations are near the action level. This can be done prospectively during survey design to

55   test the efficacy of a proposed survey design (see Chapter 4), or retrospectively during

56   interpretation of survey results to demonstrate the objectives of the survey design have been

57   achieved. The procedure for generating power curves for the Sign Test and the Wilcoxon Rank

58   Sum test are provided in Appendix I of MARSSIM. Note that the accuracy of a prospective

59   power curve depends on estimates of data variability and the planned number of measurements.

60   After the data are analyzed, the sample standard deviation provides an estimate of data

61   variability and the actual number of valid measurements are known, and these two parameters

62   are used to generate a retrospective power curve (see MARSSIM Appendix I). The consequence

63   of inadequate power is an increased Type II decision error rate. For Scenario A, this means

64   M&E that actually meet the release criteria have a higher probability of being incorrectly

65   determined not to meet the release criterion. For Scenario B, this means M&E that actually do

66   not meet the release criterion have a higher probability of being incorrectly determined to meet

67   the release criterion.

68   **6.2.2   Conduct a Preliminary Data Review**

69   A preliminary data review is performed to learn more about the structure of the data by

70   identifying patterns, relationships, or potential anomalies. The preliminary data review includes

71   reviewing quality assurance (QA) and quality control (QC) reports, performing a graphical data

72   review, and calculating basic statistical quantities.

73   6.2.2.1   Review Quality Assurance and Quality Control Reports

74   Quality assurance reports describing data collection and reporting processes provide valuable

75   information about potential problems with or anomalies in the data. EPA QA/G-9R (EPA

76   2006b) recommends a review of (1) data validation reports that document the data collection,

77   handling, analysis, reduction, and reporting procedures; (2) QC reports from laboratories or field

78   stations that document measurement system performance including data from blanks, replicates,

79   spikes, standards, and certified reference materials, or other internal QC measures; and (3)

80   technical systems reviews, performance evaluation audits, and audits of data quality including

81   data from performance evaluation measurements. EPA QA/G-9R (EPA 2006b) also suggests

82   paying particular attention to information that can be used to check assumptions made during

83   survey design using the DQO Process, especially any anomalies in recorded data, missing values,

84   deviations from SOPs, or the use of nonstandard data collection methods (e.g., new, emerging, or

85   "cutting edge" technology). Verification of instrument calibrations and review of MQOs are

86   particularly important to disposition surveys. Periodic measurements must be made to ensure the

87   measurement systems remain within acceptable calibration and control limits.

88   Quality control measurements are performed during implementation of the survey design to

89   monitor performance of the measurement methods, identify problems, and initiate corrective

90   actions when necessary. The evaluation of QC measurements used to control measurement

91   methods is distinct from the evaluation from survey results. MARLAP Section 18.3 (Evaluation

92   of Performance Indicators), Attachment 18A (Control Charts), and Attachment 18B (Statistical

93   Tests for QC Results) provide information on the evaluation of quality control measurements.

94   Reviewing QA and QC reports is the only preliminary data review performed for surveys where

95   individual measurements are not recorded (e.g., scan-only surveys with hand-held instruments).

96   This increases the importance of the QA and QC reports and should be considered during survey

97   planning to ensure data quality is adequate to meet the survey objectives.

98   6.2.2.2   Perform a Graphical Data Review

99   Preparing and evaluating graphs and other visual depictions of the data may identify trends in the

100   data that go unnoticed using purely numerical methods. The graphical data review may include

101   posting plots, frequency plots, quantile plots, or other methods for visually interpreting data.

102   General guidance on performing a graphical data review and exploratory data analysis is

103   provided in EPA QA/G-9R (EPA 2006b) and by the National Institute of Science and

104   Technology (NIST 2006). A graphical data review cannot be performed unless the measurement

105   results are recorded. Surveys where recording individual measurement results is not required

106   (e.g., scan-only surveys with hand-held instruments) do not receive a graphical data review.

107   A posting plot is simply a map of the survey unit with the data values entered at the measurement

108   locations. This type of plot potentially reveals heterogeneities in the data, especially possible

109   patches of elevated radionuclide concentrations. For a reference material survey a posting plot

110   can reveal spatial trends in background data that might affect the results of the statistical tests.

111   If the posting plot reveals systematic spatial trends in the M&E, the cause of the trends should be

112   investigated. In some cases the trends could be attributable to residual radioactivity, but they

113   may also be caused by inhomogeneities in the ambient background in the area the survey is

114   performed. EPA QA/G-9S (EPA 2006c) provides additional diagnostic tools for examining

115   spatial trends. The role of a posting plot for a conveyorized system would be a time series

116   display of the data showing any trends between adjacent batches of M&E conveyed past the

117   detector.

118   The geometric configuration of most M&E survey units composed of a few large irregularly

119   shaped pieces of M&E is transitory. The arrangement of tools and piles of scrap metal, for

120   example, changed as volumes of material were moved, or even as individual pieces were handled

121   during the survey (see Section 5.3). In these cases some identifying marks, numbers, or bar-code

122   labels should have been used to identify and track where measurements were made, at least until

123   it is determined that the M&E meet the disposition criteria. Such marking and labeling need not

124   be permanent, but may be made with materials such as chalk or removable labels.

125   A frequency plot, or histogram, is a useful tool for examining the general shape of a distribution.

126   This plot is a bar chart of the number of data points within a certain range of values. A

127   frequency plot reveals any obvious departures from symmetry, such as skewness or bimodality

128   (two peaks), in the data distributions for the M&E or reference material.

129   The presence of two peaks in the M&E data set frequency plot may indicate the presence of

130   small areas of elevated activity. In some cases it may be possible to identify an appropriate

131   background distribution within the M&E data set. This type of data interpretation generally

132   depends on site-specific considerations and should only be pursued after consultation with the

133   responsible regulatory agency.

134   The presence of two peaks in the M&E or reference material frequency plots may also indicate a

135   mixture of materials with different intrinsic radiation backgrounds. The greater variability in the

136   data caused by the presence of such a mixture reduces the power of the statistical tests. These

137   situations should be avoided whenever possible through segregation of M&E (see Section 5.4)

138   and carefully matching the reference materials to the M&E being surveyed.

139   When data are obtained from scan-only surveys incorporating data loggers, large quantities of

140   data are usually recorded. In essence, 100 percent of Class 1 M&E are measured. While the

141   survey coverage may be less than 100 percent for Class 2 and Class 3 M&E, the number of data

142   points is still likely be large. As long as there was no bias in the selection of areas that were

143   scanned, the frequency plot will be close to the population distribution of radioactivity levels in

144   the M&E. The mean and standard deviation calculated from these logged values should be very

145   close to the corresponding population values.

146   For conveyorized survey monitors, the data may be interpreted batch-by-batch as it is scanned.

147   In this case, the data treatment would be most similar to a single in situ measurement used to

148   evaluate all of the M&E. If, on the other hand, the data were logged continuously the data

149   treatment would be similar to a scan-only survey using data loggers.

150   6.2.2.3   Calculate Basic Statistical Quantities

151   Radiological survey data are usually obtained in units (e.g., counts per unit time) that have no

152   intrinsic meaning relative to the action levels. For comparison of survey data to action levels,

153   survey data from laboratory and field analyses are converted into action level units. MARSSIM

154   Section 6.6 provides guidance on data conversion. Any uncertainty associated with data

155   conversion should be included in the estimate of measurement uncertainty (see Section 5.6). For

156   surveys where individual results are not recorded (e.g., scan-only surveys with hand-held

157   instruments) the uncertainty is associated with converting the action level into the units provided

158   by the instrument in the field. Since individual results are not recorded, no statistical quantities

159   can be calculated.

160   Basic statistical quantities that should be calculated for the sample data set include the mean,

161   standard deviation, and the median. Other statistical quantities may be calculated based on the

162   survey objectives. For example, suppose the following 10 measurement results are obtained

163   from a disposition survey:

164   9.1, 10.7, 13.6, 3.4, 13.3, 7.9, 4.5, 7.7, 8.3, 10.4

165   The mean of the data (μ) is 8.89 and the standard deviation (σ) is 3.3231. [2]

166   The next 10 measurement results are from an appropriate matching reference material:

167   6.2, 13.8, 15.2, 9.3, 6.7, 4.9, 7.1, 3.6, 8.8, 8.9

168   The mean of the reference data (μ) is 8.45 and the standard deviation (σ) is 3.6713.

169   The means of the two data sets can be compared to provide a preliminary indication of the
170   survey unit status. The difference is 0.44, with the M&E being investigated having a higher
171   mean concentration. If the mean for the M&E exceeds the mean for the reference material by
172   more than the action level, the M&E clearly do not meet the disposition criterion. On the other
173   hand, if the difference between the largest M&E measurement (13.6 for this example) and the
174   smallest reference material measurement (3.6 for this example) is below the action level, the
175   M&E will pass the WRS test (Section 6.6) but will have to meet other criteria as well.

176   The value of the sample standard deviation is especially important. If the standard deviation is
177   too large compared to what was assumed for variability during development of the survey
178   design, this may indicate an insufficient number of samples were collected to achieve the desired
179   power for the statistical test. As previously mentioned, inadequate power can lead to an increase
180   in the Type II decision error rate.

181   The median is the middle value of the data set when the number of data points is odd or the mean
182   of the two middle values when the number of data points is even. A large difference between the

---

[2] Note the use of significant digits in this example. Since all of the numbers in the text are interim values in calculating the difference between two means, they are not rounded. If the mean and standard deviation values were to be reported as results they would be rounded to two significant digits because the original data is a mixture of numbers with two and three significant digits. If the data were rounded after each calculation, the difference in the rounded means appears to be 0.4 (i.e., 8.9 minus 8.5), but the actual difference is 0.44 based on the un-rounded means (i.e., 8.89 minus 8.45). This is an example of how rounding numbers too early in the process can result in additional uncertainty.

183   mean and the median indicate a potential skew in the data. This would also be evident in a

184   histogram of the data.

185   Examining other statistical quantities such as the maximum, minimum, and range may provide

186   additional useful information. For the example M&E data set the minimum is 3.4 and the

187   maximum is 13.6. The range is 13.6 - 3.4 = 10.2. The range is equal to 3.1 standard deviations

188   (i.e., 10.2/3.3). When there are thirty or fewer data points, range values greater than 4 or 5

189   standard deviations would be unusual. Thus, the range for this example data set is not unusually

190   large. The range may be greater for larger data sets.

191   **6.2.3   Select the Statistical Tests**

192   In most cases the selection of a statistical test is determined by the survey design used to collect

193   the data. The most appropriate procedure for summarizing and analyzing the data is chosen

194   based on the preliminary data review. If the preliminary data review indicates that the

195   assumptions used to develop the survey design are valid, the statistical tests and evaluation

196   methods determined should then be applied. If the assumptions used to develop the survey

197   design are determined to be invalid, it may be necessary to consult a statistician to determine the

198   most appropriate statistical test for evaluating the survey results.

199   6.2.3.1   Scan-Only Surveys

200   Scan-only surveys generate large amounts of data. Class 1 surveys measure all of the M&E.

201   When less than 100 percent of the M&E are measured (i.e., Class 2 or Class 3 surveys) the areas

202   that are measured are assumed representative of the areas that are not measured. This

203   assumption should be checked during the preliminary data review (Section 6.2.2). The

204   radionuclide concentrations or radioactivity in the areas that are not measured can be inferred

205   based on the measurement results in the areas that are measured. Data indicating this inference

206   may not be reasonable should result in re-evaluation of the survey design. For example, suppose

207   the survey design specifies that ^{137}Cs is the radionuclide of concern and scanning 50% of the

208   M&E is appropriate based on the expected distribution of radionuclide concentrations, expected

209   levels of radioactivity, and the beta-gamma emissions from the radionuclide of concern. If

210 additional historical data is found showing ^{239}Pu is also a radionuclide of concern, the survey

211 design should be re-evaluated based on the presence of an alpha emitting radionuclide as well.

212 If disposition decisions will be made for individual items or based on individual measurement

213 results, all of the results should be compared to the action level. Comparison to the action level

214 based on a detection decision or measurement (see Section 5.7) is discussed in Section 6.3.

215 Individual measurement results can be recorded for scan-only surveys. The benefit of logging

216 individual measurement results is the ability to statistically evaluate the data (e.g., calculate a

217 mean and an upper confidence limit). If disposition decisions will be made based on the mean of

218 logged data, an upper confidence level for the mean is calculated and compared to the UBGR.

219 This means that compliance with the disposition criterion can be demonstrated for the entire

220 survey unit, even if some of the results exceed the UBGR. Evaluations using the upper

221 confidence limit are discussed in Section 6.4. When less than 100% of the M&E are measured

222 (i.e., Class 2 and Class 3 surveys), the total uncertainty includes both spatial and measurement

223 uncertainty. Measuring 100% of the M&E (i.e., Class 1 survey) accounts for spatial variability,

224 but there is still an uncertainty component resulting from variability in the measurement process.

225 Conveyorized systems that continually log the survey results also generate large amounts of data.

226 An upper confidence level for the mean can be used for the evaluation of data from these types

227 of systems (see Section 6.4) in the same manner as logged scan data. Conveyorized systems that

228 operate in a batch mode are essentially treated as single in situ measurements of small batches of

229 M&E. The results generated by these types of systems are evaluated as a series of comparisons

230 to the UBGR; using detection decisions based on the MDC (see Section 6.3).

231 6.2.3.2   In Situ Surveys

232 In situ surveys may consist of a series of isolated measurements covering all or part of the M&E

233 (i.e., MARSSIM-type survey design), a series of measurements with overlapping fields of view

234 incorporating all (Class 1) or a portion (Class 2 or Class 3) of the M&E, or a single measurement

235 incorporating all of the M&E (see Section 4.4.2). Different assumptions are used to design each

236 of these types of in situ surveys, and different methods are used to evaluate the results of these

237 surveys.

238   Similar to scan-only surveys, if disposition decisions will be made for individual items or based

239   on individual measurement results, all of the results should be compared to the action level.

240   Comparison to the action level based on a detection decision (see Section 5.7) is discussed in

241   Section 6.3. Unlike scan-only surveys, in situ surveys are likely based on a limited number of

242   data points. To perform in situ measurements, assumptions were made about the distribution of

243   radioactivity within the volume of M&E being measured. These assumptions are inherent in the

244   calibration of in situ measurement systems and the validity of these assumptions determines the

245   appropriateness of the measurement. It is important to account for uncertainty in these

246   assumptions when calculating the MDC and to evaluate these assumptions using QC

247   measurements performed during the survey. If there is uncertainty about the true MDC or

248   critical value, use conservative values for the efficiency as described in Chapter 5.

249   6.2.3.3   MARSSIM-Type Survey Designs

250   MARSSIM-type survey designs are generally used when instrumentation for scan-only or in situ

251   measurement surveys does not provide sufficient sensitivity (e.g., the MDC is greater than the

252   UBGR). A statistically based number of measurements is used to provide an estimate of the

253   mean activity in each survey unit, and scanning is used to identify small areas of elevated

254   activity between sample locations.

255   The number of measurements is determined by the statistical test. In most cases the statistical

256   tests used in MARSSIM are appropriate for Scenario A. The criteria for choosing between the

257   Sign test and the Wilcoxon Rank Sum (WRS) test are described in MARSSIM Section 8.2.3. In

258   general, when the radionuclide is not present in background (or its background concentration is

259   negligible compared to the action level) and radionuclide-specific measurements are made, the

260   Sign test (Section 6.5) is used. Otherwise, the WRS (Section 6.6) test should be used. The Sign

261   test is designed to detect whether there is radioactivity in the M&E above the action level. The

262   WRS test is used to compare measurements of the M&E to measurements performed on the

263   reference material.

264   When Scenario B is used, the statistical tests described in NUREG-1505 (NRC 1998a) are

265   generally used. The Sign test and the WRS test are still used, but the application of the test is

266   adjusted to account for the difference in the null hypothesis. When using Scenario B, there is a

267   potential for the WRS test to miss non-uniform radioactivity (i.e., slightly elevated radionuclide

268   concentrations or levels of radioactivity over a portion of the survey unit). Randomization of the

269   M&E through mixing or homogenization can eliminate this possibility. If randomization is not

270   practical, the Quantile test (Section 6.7) should be used to evaluate survey units when the WRS

271   test fails to reject the null hypothesis.

272   The results of scanning measurements performed as part of a MARSSIM-type survey are

273   evaluated using the elevated measurement comparison (EMC). The EMC is simply a

274   comparison to an action level (see Section 6.3). The action level used for the EMC is the action

275   level for small areas of elevated activity. If there is no action level for elevated activity, the

276   scanning results are compared to the action level for the mean activity in the survey unit.

277   Additional information on the EMC is available in MARSSIM Section 8.5.1 and NUREG-1505

278   Chapter 8 (NRC 1998a).

279   **6.2.4   Verify the Assumptions of the Tests**

280   An evaluation to determine the data are consistent with the underlying assumptions of the

281   statistical tests helps to validate the use of a particular test. One may also determine that certain

282   departures from these assumptions are acceptable when given the actual data and other

283   information about the project. The nonparametric tests described in this chapter assume that the

284   data from the M&E or the reference material consist of independent measurements from each

285   distribution. The primary issue associated with the evaluation of scan-only and single in situ

286   measurement survey data is the MDC or MQC as discussed in Section 6.2.1.

287   Asymmetry in the data can be identified using a histogram or a Quantile plot. Information on

288   histograms and Quantile plots is provided in MARSSIM Appendix I and NUREG-1505 Section

289   4.2.2 (NRC 1998a). As discussed in Section 6.2.2.3, data transformations can sometimes be used

290   to minimize the effects of asymmetry.

291   One of the primary advantages to using the nonparametric tests is that they involve fewer

292   assumptions about the data than their parametric counterparts. If parametric tests are used (e.g.,

293   Student's t test) any additional assumptions made in using these tests should be verified (e.g.,

294   testing for normality). These issues are discussed in detail in EPA QA/G-9S (EPA 2006c).

295   One of the more important assumptions made in the survey design is that the number of

296   measurements is sufficient to achieve the DQOs set for the Type I (α) and Type II (β) decision

297   error rates. Verification of the power of the statistical tests (1-β) may be of particular interest.

298   Methods for assessing power are discussed in Appendix I.9 of MARSSIM. If there is not

299   reasonable assurance the DQOs have been achieved, additional investigations including

300   repeating the survey may be needed. The planning team can develop survey designs cautiously

301   to avoid unnecessary and potentially costly decision errors by:

302   • Estimating the potential data variability conservatively,

303   • Taking more measurements than suggested by the DQO process, and

304   • Estimating the MDCs conservatively.

305   In the absence of other data, each of these estimates could be multiplied by a safety factor of 1.2

306   (i.e., increase the estimate by 20%). Examples of assumptions and possible methods for

307   evaluating and verifying these assumptions are summarized in Table 6.1.

308   Verification of scan-only and in situ survey results focuses on the estimates of the MDC and

309   MQC values used to design the survey. If the assumptions used to estimate these values are

310   incorrect, the survey design may be invalid.

311   The first step in evaluating the MDC and MQC is to review the assumptions used to develop

312   these values. In general, the key assumptions are made in determining the source and detector

313   efficiencies. QA and QC reports should be reviewed to evaluate measurement performance (e.g.,

314   scan speed, source geometry, distance from M&E to the detector, non-uniform response of large

315   area detectors). The description of the M&E from the IA should be compared to the assumptions

316   used to develop the efficiency.

317   In some cases it may be possible to compare the survey results of multiple measurement

318   techniques. For example, if there are multiple radiations associated with the M&E it may be

319   possible to compare gamma measurement results to alpha or beta measurement results to verify

320   the survey results. Direct measurements may provide more quantitative results for areas of

321   elevated activity identified during scan-only surveys.

322   **Table 6.1 Issues and Assumptions Underlying the Evaluation Method**

Evaluation Method	Issue	Verification Method	Survey Type
Compare single measurements to a limit (see Section 6.3)	Verify the MDC and Measurement Uncertainty	Review the MDC Review QA/QC Reports Review IA and DQOs	Scan-Only In situ
Compare an upper confidence limit for the mean to a limit (see Section 6.4)	Verify the MQC and Measurement Uncertainty	Review the Measurement Uncertainty Review QA/QC Reports Review IA and DQOs	Scan Only In situ
Statistical Tests (see Sections 6.5, 6.6, and 6.7)	Verify the Assumptions of the Statistical Test (e.g., spatial independence, symmetry, data variance, power)	Preliminary Data Review (e.g., posting plot, histogram, summary statistics, power curve)	MARSSIM-Type Survey

323   It may be possible to use an entirely different survey method to provide information to support

324   verification of assumptions used to design a survey. For example, smears or surface scrapings

325   can be used to verify the presence of radionuclides or radioactivity on the surface.[3] In situ

326   measurements or sample collection and analysis may be used to verify the results of scan-only

327   survey designs. Care must be taken to ensure comparability of survey methods before evaluating

328   the results to avoid generating conflicting results. For example, consider an in situ survey used

329   to demonstrate the mean activity is less than the action level. A scan-only survey method is used

330   to verify the results and identifies an area of elevated activity. This discrepancy in results

331   warrants additional investigation of the small area of elevated activity. The additional

332   investigation should determine if the activity in this area actually causes the mean activity to

333   exceed the disposition criterion.

_____

[3] This smear procedure does not rule out additional volumetric activity.

334 **6.2.5   Draw Conclusions from the Data**

335   The types of measurements performed on M&E are:

336   • Scans,

337   • In situ or direct measurements at discrete locations, and

338   • Samples collected at discrete locations.

339   Specific details for conducting the Sign test and the WRS tests are provided in Sections 6.5 and
340   6.6, respectively. When the data clearly show that the M&E meets or exceeds the disposition
341   criterion, the result is often obvious without performing the formal statistical analysis. This is
342   the expected outcome for Class 2 and Class 3 surveys. Table 6.2 summarizes examples of
343   circumstances leading to specific conclusions based on a simple examination of the data.

344   **6.3   Compare Results to the UBGR**

345   When disposition decisions will be made about individual items, or decisions will be based on
346   individual measurement results, each result (plus or minus a multiple of its combined standard
347   uncertainty) will be compared to the action level (see MARLAP Appendix C.4). In practice, this
348   means that any result that exceeds the critical value (S_C, see Section 5.7.1) when the minimum
349   detectable level (S_D, see Section 5.7.2) equals the UBGR provides evidence that the result
350   exceeds the UBGR.

351   For Scenario A, if all the results are less than the action level, then the mean and the maximum
352   activity must also be below the action level. Thus, the radionuclide concentrations or levels of
353   radioactivity associated with the M&E demonstrate compliance with the disposition criterion.
354   For Scenario B when the action level is not zero or background, all of the results must be below
355   the critical value corresponding to the MDC set equal to the UBGR. If the action level is zero or
356   background, Scenario B must be used and any indication of the presence of radionuclide
357   concentrations or radioactivity above background (i.e., above the discrimination level) would
358   result in rejecting the null hypothesis. For this situation, any measurement result exceeding the
359   critical value corresponding to the required MDC indicates the potential presence of
360   radionuclides or radioactivity above background. This applies to single in situ measurements as
361   well as series of in situ measurements.

362 **Table 6.2 Summary of Evaluation Methods and Statistical Tests**

Evaluation Method or Statistical Test	Survey Result	Conclusion
Comparison to a Limit (AL=0) *Scenario B only* *Results may or may not be recorded* *Scan-only or In situ surveys*	All measurements less than the critical value corresponding to the MDC (e.g., does not exceed alarm set point)	M&E meet the disposition criterion
	Any measurement exceeds the critical value corresponding to the MDC	M&E do not meet the disposition criterion
Comparison to a Limit (AL≠0) *Scenario A or B* *Results not recorded* *Scan-only or In situ surveys*	All measurements less than the critical value corresponding to the UBGR	M&E meet the disposition criterion
	Any measurement exceeds the critical value corresponding to the UBGR	M&E do not meet the disposition criterion
Comparison to Upper Confidence Limit *Scenario A or B* *Results must be recorded* *Scan-only or In situ surveys*	Upper confidence limit less than UBGR	M&E meet the disposition criterion
	Upper confidence limit greater than UBGR	M&E do not meet the disposition criterion
Sign Test *Radionuclide not in background* *Nuclide-specific measurements* *Scenario A or B* *MARSSIM-type surveys*	All measurements less than the action level	M&E meet the disposition criterion
	Mean greater then the action level	M&E do not meet the disposition criterion
	Any measurement greater than the action level and the mean less than the action level	Conduct Sign test (and elevated measurement comparison, if necessary)
Wilcoxon Rank Sum (WRS) Test *Radionuclide in background* *Nuclide non-specific measurements* *Scenario A or B* *MARSSIM-type surveys*	Difference between maximum survey unit measurement and minimum reference area measurement is less than the UBGR	M&E meet the disposition criterion
	Difference of survey unit mean and reference area mean is greater than the action level	M&E do not meet the disposition criterion

Evaluation Method or Statistical Test	Survey Result	Conclusion
	Difference between any survey unit measurement and any reference area measurement greater than the action level or the difference of survey unit mean and reference area mean is less than the action level	Conduct WRS test (and elevated measurement comparison, if necessary)
Quantile Test *Test for non-uniform radioactivity* *Combine with WRS test* *Scenario B only* *MARSSIM-type surveys*	Difference between maximum survey unit measurement and minimum reference area measurement is less than the UBGR	M&E meet the disposition criterion
	Difference of survey unit mean and reference area mean is greater than the action level	M&E do not meet the disposition criterion
	Difference between any survey unit measurement and any reference area measurement greater than the action level or the difference of survey unit mean and reference area mean is less than the action level	Conduct Quantile test (and elevated measurement comparison, if necessary)

363 If there is an action level based on small areas of elevated activity or the maximum allowable
364 value, the individual results can be compared directly to the action level. This applies primarily
365 to the evaluation of scanning results for MARSSIM-type surveys (i.e., the EMC), but may be
366 applied to scan-only survey data as well.

367 **6.4 Compare Results Using an Upper Confidence Limit**

368 When disposition decisions are made about the estimated mean of a sampled population, the
369 assessment of the survey results is accomplished by comparing an upper confidence limit for the

370   mean to the UBGR. For scan-only surveys where there are a large number of data points, a

371   simple comparison of the mean activity to the UBGR may be sufficient.[4]

372   If individual scan-only survey results are recorded, a non-parametric confidence interval can be

373   used to evaluate the results of the disposition survey. Similarly, a confidence interval can be

374   used to evaluate a series of in situ measurements with overlapping fields of view. A one-tailed

375   version of Chebyshev's inequality or EPA's ProUCL software can be used to evaluate the

376   probability of exceeding the UBGR (i.e., using an upper confidence limit). The use of an upper

377   confidence limit applies to both Scenario A (where the UBGR equals the action level) and

378   Scenario B when the action level is not zero or background (where the UBGR equals the

379   discrimination limit). Comparison to an upper confidence level should not be used for Scenario

380   B when the action level equals zero or background.[5]

381   If all of the survey results are less than the UBGR, the mean must also be less than the UBGR.

382   In this situation it is not necessary to calculate the upper confidence limit to show that the survey

383   results demonstrate compliance with the disposition criterion.

---

[4] The calculation of the upper confidence limit is based partially on the number of measurements. When the number of measurements gets very large, i.e., close to 100% of the M&E have been measured, the estimates for the mean and the standard deviation get close to the actual, but unknown, values. In certain situations (e.g., 100% measurement with a documented measurement method) the planning team may assume that the estimate of the mean based on the survey results is close enough to the actual, but unknown, mean of the population to be compared directly to the UBGR.

[5] By convention, any result exceeding the critical level is not likely to be zero or background since $1-\alpha$ (i.e., 95% for the MDC) of the background distribution is below the critical level. This means that the null hypothesis stating the radionuclide concentration or level of radioactivity in the M&E is zero or consistent with background should be rejected.

384   **6.4.1   Calculate the Upper Confidence Limit**

385   Chebyshev's inequality calculates the probability that the absolute value of the difference of the

386   true but unknown mean of the population and a random number from the data set is at least a

387   specified value. That is, given a specified positive number (n), a mean (μ), and a random number

388   from the data set (r), then the probability that $|\mu\text{-}r|$ is greater than or equal to n is equal to α. In

389   addition, a one-tailed version of the inequality can be used to calculate an upper confidence limit

390   (UCL) for a data set that is independent of the data distribution (i.e., there is no requirement to

391   verify the data are from a normal, lognormal, or any other specified kind of distribution) by

392   letting the inequality equal the UCL, as described in the following steps.

393   1. Calculate the mean (μ) and standard deviation (σ) of the number of results (n) in the data

394   set.

395   2. For Scenario A, retrieve the Type I error rate (α) used to design the survey.

396   3. Using Chebyshev's inequality, calculate the maximum UCL using equation 6-1:

397
$$\text{UCL} = \mu + \sqrt{\frac{\sigma^2}{n\alpha} - \frac{\sigma^2}{n}}$$
(6-1)

398   4. For Scenario B, substitute the Type II error rate (β) used to design the survey for α in

399   Equation 6-1.

400   5. If the maximum UCL is less than the UBGR, the survey demonstrates compliance with

401   the disposition criterion (i.e., reject the null hypothesis for Scenario A or fail to reject the

402   null hypothesis for Scenario B).

403   Chebyshev's inequality must be used with caution when there are very few points in the data set.

404   This is because the population mean and standard deviation in the Chebyshev formula are being

405   estimated by the sample mean and sample standard deviation. In a small data set from a highly

406   skewed distribution, the sample mean and sample standard deviation may be underestimated if

407   the high concentration but low probability portion of the distribution is not captured in the

408   sample data set. EPA has issued guidance on calculating upper confidence limits for exposure

409   point concentrations (EPA 2002b)[6]. Software for implementing EPA's guidance is available

410   (EPA 2006d).

411   **6.4.2   Upper Confidence Limit Example: Class 1 Concrete Rubble**

412   This example illustrates the survey design for concrete rubble using 3 in x 3 in NaI(Tl) detectors

413   mounted on a conveyorized survey system to measure ^{137}Cs. A pile of concrete rubble was

414   loaded on the conveyor and passed beneath the detectors at a pre-determined speed. Each one-

415   second count recorded by a detector corresponds to approximately 9,800 cm^3 of concrete rubble

416   (i.e., a 5 cm thick disk with a 50 cm diameter). The following information was used to design

417   the survey:

418   • The selected disposition option was clearance, using Scenario A with the null

419   hypothesis that the residual radioactivity exceeds the action level.

420   • The IA indicated the concrete was potentially volumetrically contaminated prior to

421   being converted to rubble.

422   • The concrete rubble had a maximum particle dimension of less than 0.5 cm.

423   • The average background count rate was estimated to be 38,000 cpm based on

424   preliminary surveys of non-impacted concrete, and was used for the LBGR.

425   • The action level was set at 20,000 cpm above the average background count rate, so

426   the UBGR was set at 58,000 cpm.

427   • The estimated standard deviation of background count rate is 2,500 cpm based on

428   preliminary survey data.

429   • The Type I decision error rate was set at 0.10, or 10%.

430   The survey consisted of 9,616 one-second measurements that were recorded using a data logger.

431   The mean count rate for the survey was 39,252 cpm, with a standard deviation of 5,465 cpm.

---

[6] In MARSAME, exposure point concentration is used to mean a conservative estimate of the mean radionuclide concentration(s) in or on M&E.

432   The standard deviation of the mean was $5465/\sqrt{9616} = 55.7$. As noted earlier, with such a large

433   data set, one can expect that the sample mean and standard deviation should be fairly close to

434   their population values. Note that the estimated coefficient of variation of the mean is

435     $55.7/39,252 = 0.14\%$. The minimum count rate was 30,080 cpm, and the maximum count rate

436   was 72,805 cpm. Note that although the mean concentration is well below the action level, there

437   are data points that exceed the action level. Thus, a test against an UCL for the mean is

438   warranted. Figure 6.1 shows a frequency plot of the survey results.

439
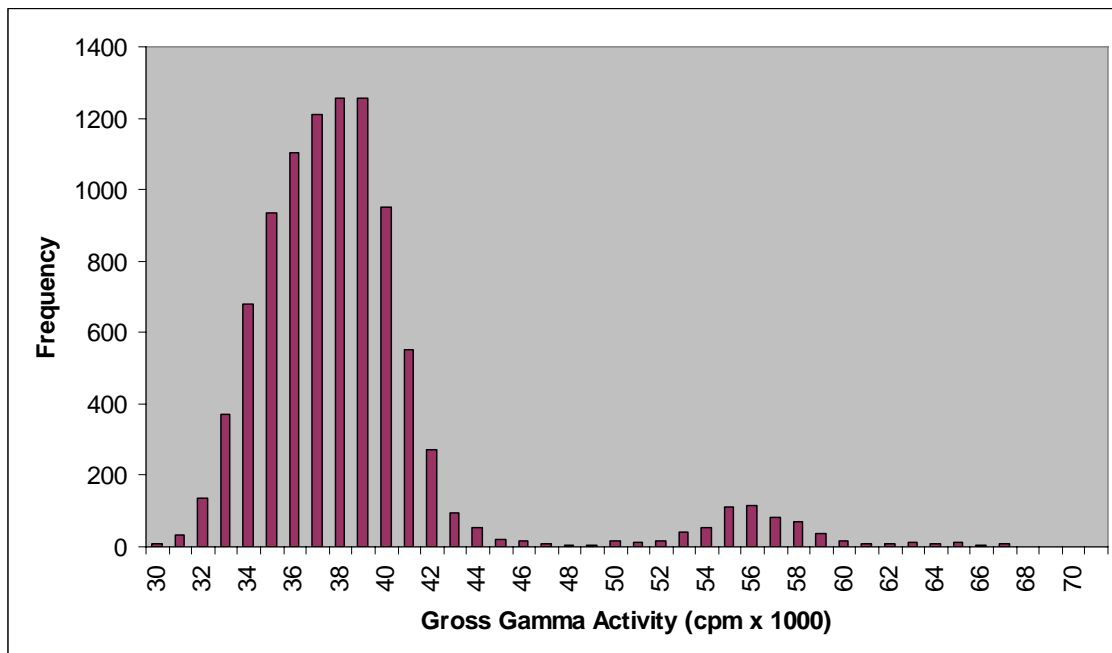
440   **Figure 6.1 Frequency Plot of Concrete Rubble Data**

441   If the sample size were small, however, the upper part of the bimodal distribution could be

442   missed and the Chebyshev UCL could be underestimated. In this case, with a sample size of

443   9,616, the upper confidence level was calculated as described in Section 6.4.1.

444               $$UCL = 39,252 + \sqrt{\frac{(5,465)^2}{(0.10)(9,616)} + \frac{(5,465)^2}{(9,616)}} = 39,474 \text{ cpm}$$

445   The upper confidence level of 39,474 cpm is much less than the action level of 58,000 cpm. The

446   null hypothesis that the level of radioactivity exceeds the disposition criterion is rejected. The

447   EPA ProUCL software was also applied to these data and the results are shown in Figure 6.2.

448   The software has failed to find a good fit to the data for normal, lognormal or gamma

449   distributions, which is hardly surprising given the bimodal nature of the data. The

450   recommendation is that either a Student's t or a modified Student's t 95% UCL be used. These

451   are both listed as about 39,343. These are lower than the 90% Chebyshev UCL of 39,474 used

452   above, but that would not change the conclusion. A 95% Chebyshev UCL calculated according

453   to Section 6.4.1 would have been 39,574. Note that the 95% Chebyshev UCL calculated by

454   ProUCL, rounded to the nearest count, is slightly different, 39,495, because of the way that the

455   sample mean and standard deviation are estimated before entering them in the Chebyshev

456   formula. The ProUCL User's Manual can be consulted for details. However, with the number

457   of data points at hand, there is little difference among any of the methods for computing an UCL.

458   **6.5   Sign Test**

459   The Sign test is used to compare the measurement results from each survey unit with the

460   applicable disposition criterion. The Sign test can be applied to either Scenario A or Scenario B.

461   The Sign test should only be used if the radionuclide being measured is not present in

462   background. The Sign test may also be used if the radionuclide being measured is present at

463   such a small fraction of the action level as to be considered insignificant. Otherwise, the WRS

464   test described in Section 6.6 should be applied. Additional information on the Sign test can be

465   found in Section 8.3 of MARSSIM and Chapter 5 of NUREG 1505 (NRC 1998a).

UCL Statistics for CPM

	A	B	C	D	E	F	G	H	I
1	Data File					Variable:	CPM		
2									
3		Raw Statistics				Normal Distribution Test			
4	Number of Valid Samples			9616	Lilliefors Test Statisitic				0.2044466
5	Number of Unique Samples			6441	Lilliefors 5% Critical Value				0.0090352
6	Minimum			30080	Data not normal at 5% significance level				
7	Maximum			72805					
8	Mean			39251.847		95% UCL (Assuming Normal Distribution)			
9	Median			38267	Student's-t UCL				39343.497
10	Standard Deviation			5465.0563					
11	Variance			29866840		Gamma Distribution Test			
12	Coefficient of Variation			0.1392306	A-D Test Statistic				585.96505
13	Skewness			2.4504964	A-D 5% Critical Value				0.7522512
14					K-S Test Statistic				0.1788147
15		Gamma Statistics			K-S 5% Critical Value				0.01814
16	k hat			61.605609	Data do not follow gamma distribution				
17	k star (bias corrected)			61.586458	at 5% significance level				
18	Theta hat			637.14729					
19	Theta star			637.34542		95% UCLs (Assuming Gamma Distribution)			
20	nu hat			1184799.1	Approximate Gamma UCL				39335.904
21	nu star			1184430.8	Adjusted Gamma UCL				39335.917
22	Approx.Chi Square Value (.05)			1181899.7					
23	Adjusted Level of Significance			0.049975		Lognormal Distribution Test			
24	Adjusted Chi Square Value			1181899.4	Lilliefors Test Statisitic				0.1656286
25					Lilliefors 5% Critical Value				0.0090352
26		Log-transformed Statistics			Data not lognormal at 5% significance level				
27	Minimum of log data			10.311616					
28	Maximum of log data			11.19554		95% UCLs (Assuming Lognormal Distribution)			
29	Mean of log data			10.569616	95% H-UCL				N/A
30	Standard Deviation of log data			0.1224578	95% Chebyshev (MVUE) UCL				39441.02
31	Variance of log data			0.0149959	97.5% Chebyshev (MVUE) UCL				39533.758
32					99% Chebyshev (MVUE) UCL				39715.923
33									
34						95% Non-parametric UCLs			
35					CLT UCL				39343.516
36					Adj-CLT UCL (Adjusted for skewness)				39345.004
37					Mod-t UCL (Adjusted for skewness)				39343.729
38					Jackknife UCL				39343.497
39					Standard Bootstrap UCL				39341.993
40					Bootstrap-t UCL				39342.895
41		RECOMMENDATION			Hall's Bootstrap UCL				39343.322
42		Data are Non-parametric (0.05)			Percentile Bootstrap UCL				39337.588
43					BCA Bootstrap UCL				39344.539
44		Use Student's-t UCL			95% Chebyshev (Mean, Sd) UCL				39494.773
45		or Modified-t UCL			97.5% Chebyshev (Mean, Sd) UCL				39599.887
46					99% Chebyshev (Mean, Sd) UCL				39806.364

466

467   **Figure 6.2 Output from ProUCL Software for the Sample Data Set**

468   **6.5.1   Apply the Sign Test to Scenario A**

469   The Sign test is applied to Scenario A by counting the number of measurements from each

470   survey unit that are less than the action level (i.e., UBGR). Each result is subtracted from the

471   action level ($AL - X_i$), and the number of positive values is summed. The result is the test

472   statistic $S+$. Discard any measurement that is exactly equal to the action level and reduce the

473   sample size, N, by the number of such measurements. The value of $S+$ is compared to the

474   critical values in A.3. If $S+$ is greater than the critical value (k) in the table, the null hypothesis

475   is rejected.

476   **6.5.2   Apply the Sign Test to Scenario B**

477   The Sign test is applied to Scenario B in a manner similar to that used for Scenario A. However,

478   for Scenario B the action level (i.e., LBGR) is subtracted from each result ($X_i - AL$), and the

479   number of positive values is summed. The result is the test statistic $S+$. Discard any

480   measurement that is exactly equal to the action level and reduce the sample size, N, by the

481   number of such measurements. The value of $S+$ is compared to the critical values in Table A.3.

482   If $S+$ is greater than the critical value (k) in the table, the null hypothesis is rejected.

483   **6.5.3   Sign Test Example: Class 1 Copper Pipes**

484   This example illustrates the disposition survey design for copper pipe sections using a gas-flow

485   proportional counter to measure ^{239}Pu. Since the alpha background on the copper material is

486   essentially zero, it was decided the Sign test would be used to determine whether the material

487   meets the disposition criterion. The sample size was determined using the DQO Process and

488   inputs such as the disposition option, action level, expected standard deviation of the

489   measurement results, and the acceptable probability of making Type I and Type II decision

490   errors.

491   The following inputs were used to develop the survey design:

492   •   The selected disposition option was clearance.

493   •   The survey was designed using Scenario A, with the null hypothesis that the residual
494   radioactivity exceeds the action level.

495   •   The IA indicated that the inside surfaces of the pipes potentially came in contact with
496   liquids containing ^{239}Pu, but the outside surfaces were non-impacted.

497   •   The gross activity action level was 100 dpm/100 cm^2. When converted to cpm the
498   gross activity action level was 10 cpm (i.e., total efficiency = 0.10 counts per
499   disintegration).

500   •   The LBGR (i.e., the DL) was set at the expected activity level on the copper pipe
501   sections (i.e., 5 net cpm - the same as the gross mean for an alpha background of
502   zero).

503   •   The standard deviation for the measurements was estimated at 2 cpm.

504   •   The relative shift was calculated as (10-5)/2 = 2.5.

505   •   The Type I and Type II decision error rates were both set at 0.05.

506   Table A.2a indicates the number of measurements estimated to be needed for the Sign test, N, is
507   15 (α=0.05, β=0.05, and Δ/σ=2.5). Therefore 15 surface activity measurements were randomly
508   collected from the inside surfaces of the copper pipe sections. Survey results are shown in
509   Table 6.3.

510   The surface activity values in Table 6.3 were determined by dividing the measured cpm by the
511   total efficiency (0.10). No probe area correction was necessary. The mean count rate was 5
512   cpm, compared to the estimate of 5 cpm used for the LBGR, and the median was 4 cpm. The
513   standard deviation was 4 cpm, which was higher than the value of 2 used to develop the survey

514   design.[7] Thus, the power of the test is lower than planned. With the actual value of the relative

515   shift (10-5)/4=1.2, 23 measurements should have been collected.

516   **Table 6.3 Sign Test Example Data**

Surface Concentration (cpm/100 cm^2)	Surface Concentration (dpm/100 cm^2)	< Action Level?
4	40	Yes
3	30	Yes
11	110	No
1	10	Yes
1	10	Yes
4	40	Yes
6	60	Yes
3	30	Yes
9	90	Yes
6	60	Yes
14	140	No
1	10	Yes
4	40	Yes
10	100	No
2	20	Yes
Number of measurements less than the action level (S+) = 12		

517   With the 15 measurements collected, the actual Type II decision error rate was between 0.10 and

518   0.25 (the closest entries in Table A.2a are for α=0.05, β=0.10, and Δ/σ=1.2 with N=18, and

519   α=0.05, β=0.25, and Δ/σ=1.2 with N=12). Three measurements exceeded the action level. The

520   portion of the material associated with these measurements merits further investigation.

---

[7] Values are reported to one significant figure based on the data in Table 6.3. Interim calculations generally carry extra figures, so rounding to the appropriate number of significant figures only occurs for the final calculation. Rounding results too soon in the calculation may result in unnecessarily deleting individual results (i.e., when the result is exactly equal to the UBGR) resulting in lower statistical power.

521   The value of $S+$, 12, was compared to the appropriate critical value, q, in Table A.3. In this case,

522   for $N=15$ and $\alpha=0.05$, the critical value is 11. Since $S+$ exceeds q, reject the null hypothesis that

523   the survey unit exceeds the action level. In this case, the slight loss of power attributable to

524   underestimating the standard deviation did not affect the result. Pending the outcome of the

525   investigation of the three elevated measurements, this survey unit has satisfied the disposition

526   criteria established for clearance.

6.6   Wilcoxon Rank Sum Test

527

528   The Wilcoxon Rank Sum (WRS) test is used to compare each material survey unit with an

529   appropriately chosen reference material. Each reference material should be selected on the basis

530   of its similarity to the survey unit material, as discussed in Section 3.9. The WRS test can be

531   applied to either Scenario A or Scenario B. Further information on the WRS test can be found in

532   Section 8.4 of MARSSIM and Chapter 6 of NUREG- 1505 (NRC1998a).

6.6.1   Apply the WRS Test to Scenario A

533

534   The WRS test is applied to Scenario A as outlined in the following steps and further illustrated

535   by the example in Section 6.6.2.

536   1. Obtain the adjusted reference material measurements, Z_i, by adding the action level to

537   each reference material measurement, X_i. $Z_i = X_i +$ AL.

538   2. The m adjusted reference sample measurements, Z_i, from the reference material and the n

539   sample measurements, Y_i, from the survey unit are pooled and ranked in order of

540   increasing size from 1 to N, where $N = m+n$.

541   3. If several measurements are tied (i.e., have the same value), they are all assigned the

542   mean rank of that group of tied measurements.

543   4. If there are t "less than" values, they are all given the mean of the ranks from 1 to t.

544   Therefore, they are all assigned the rank $t(t+1)/(2\ t) = (t+1)/2$, which is the mean of the

545   first t integers. If there is more than one MDC,[8] all observations below the largest MDC

546   should be treated as "less than" values. If more than 40% of the data from either the

547   reference material or the survey unit are reported as less than detectable, the WRS test

548   *cannot* be used.

549   5. The sum of all the ranks, which is the sum of the first N positive integers, is $N(N+1)/2$,

550   which equals W_r added to W_s. Thus, one needs only to sum the ranks of the either the

551   adjusted reference measurements (W_r) or the sum of the ranks of the sample

552   measurements (W_s).

553   6. Compare W_r with the critical value (q) given in Table A.4 for the appropriate values of n,

554   m, and α. If W_r is greater than the tabulated value for q, reject the hypothesis that the

555   survey unit exceeds the disposition criterion.

556 **6.6.2 Apply the WRS Test to Scenario B**

557 The WRS test is applied to Scenario B as outlined in the following steps:

558   1. Obtain the adjusted survey unit measurements, Z_i, by subtracting the LBGR from each

559   survey unit measurement, Y_i. $Z_i = Y_i$ - LBGR.

560   2. The n adjusted survey unit measurements, Z_i, and the m reference material measurements,

561   X_i, are pooled and ranked in order of increasing size from 1 to N, where $N = m+n$.

562   3. If several measurements are tied (i.e., have the same value), they are all assigned the

563   mean rank of that group of tied measurements.

---

[8] Examples of situations where there could be more than one MDC include using multiple laboratories to perform sample analyses and using different instruments with different backgrounds and different efficiencies to perform measurements.

564   4. If there are t "less than" values, they are all given the mean of the ranks from 1 to t.

565   Therefore, they are all assigned the rank $t(t+1)/(2\ t) = (t+1)/2$, which is the mean of the

566   first t integers. If there is more than one MDC, all observations below the largest MDC

567   should be treated as "less than" values. If more than 40% of the data from either the

568   reference material or the survey unit are reported as less than detectable, the WRS test

569   *cannot* be used.

570   5. Sum the ranks of the adjusted measurements from the survey unit, W_s. The sum of all the

571   ranks, which is the sum of the first N positive integers, is $N(N+1)/2$, which equals W_r

572   added to W_s. Thus, one needs only to sum the ranks of the either the adjusted reference

573   measurements (W_r) or the sum of the ranks of the sample measurements (W_s).

574   6. Compare W_s with the critical value (q) given in Table A.4 for the appropriate values of n,

575   m, and α. (NOTE: When using this table for Scenario B, the roles of m and n are

576   reversed. If the Quantile test is being used in addition to the WRS test, then $\alpha/2$ should

577   be used rather than α.) If W_s is greater than the tabulated value for q, reject the

578   hypothesis that the difference in the median concentration between the survey unit and

579   the reference area is less than the LBGR.

580 **6.6.3 WRS Test Scenario A Example: Class 2 Metal Ductwork**

581 This example illustrates the use of the WRS test for releasing Class 2 metal ductwork. Assume

582 that a gas-flow proportional detector was used to make gross (non-radionuclide-specific) surface

583 activity measurements.

584 The DQOs from this survey unit include $\alpha = 0.05$ and $\beta = 0.05$, and the action level converted to

585 units of gross cpm is 2,300 cpm, which is the UBGR. In this case, the WRS test was used

586 because the estimated background level (2,100 cpm) was large compared to the action level. The

587 estimated standard deviation of the measurements, σ, was 375 cpm. The estimated added

588 activity level was 800 cpm; the LBGR was set at this value, and represents the DL. The relative

589 shift was calculated as Δ/σ, which is (action level – LBGR)/σ, which equals 4.

590   The sample size needed for the WRS test can be found in Table A.2b for these DQOs. The result

591   is nine measurements in each survey unit and nine in each reference material ($\alpha = 0.05$, and $\beta =$

592   0.05, and $\Delta/\sigma = 4$). The ductwork was laid flat onto a prepared grid, and the nine measurements

593   needed in the survey unit were made using a random-start triangular grid pattern. For the

594   reference materials, the measurement locations were chosen randomly on a suitable batch of

595   material. Table 6.4 lists the gross count rate data obtained.

596   **Table 6.4 Scenario A WRS Test Example Data**

Data (cpm)	Area	Adjusted Data	Ranks	Reference Material Ranks
2180	R	4480	15	15
2398	R	4698	16	16
2779	R	5079	18	18
1427	R	3727	10	10
2738	R	5038	17	17
2024	R	4324	13	13
1561	R	3861	11	11
1991	R	4291	12	12
2073	R	4373	14	14
2039	S	2039	3	0
3061	S	3061	8	0
3243	S	3243	9	0
2456	S	2456	7	0
2115	S	2115	4	0
1874	S	1874	2	0
1703	S	1703	1	0
2388	S	2388	6	0
2159	S	2159	5	0
		Sum =	171	126

597 In the "Area" column, the code "R" denotes a reference material measurement and "S" denotes a

598 survey unit measurement. The adjusted data were obtained by adding the action level to the

599 reference material measurements (see Section 6.6.1, Step 1). The ranks of the data range from 1

600 to 18, since there are a total of 9+9 measurements (see Section 6.6.1, Step 2). Note that the sum

601 of all of the ranks is still $18(18+1)/2 = 171$. Checking this value with the formula in Step 5 of

602 Section 6.6.1 is recommended to guard against errors in the rankings.

603 The total of the ranks belonging to the reference material measurements is 126. This is

604 compared with the entry for the critical value of 104 in Table A.4 for $\alpha = 0.05$, with $n = 9$ and

605 $m = 9$. Since the sum of the reference material ranks is greater than the critical value, the null

606 hypothesis (i.e., that the mean survey unit concentration exceeds the action level) is rejected, and

607 the ductwork is released.

608 This conclusion can be reached quickly by noting the difference between the largest survey unit

609 measurement (3,243 cpm) and the smallest reference area measurement (1,427 cpm). This

610 difference ($3,243 – 1,427 = 1,816$ cpm) is less than the action level of 2,300 cpm. Since the

611 largest possible difference is less than the action level, the mean difference must also be less than

612 the action level.

613 **6.6.4 WRS Test Scenario B Example: Class 2 Metal Ductwork**

614 This example illustrates the use of the Scenario B WRS test for releasing Class 2 metal

615 ductwork, using the same data as in Section 6.6.3. The null hypothesis for Scenario B is that

616 there is no detectable radioactivity above background.

617 In this case the action level was set at no radioactivity detectable above the estimated

618 background level (2,100 cpm). The LBGR is equal to the action level, and is set to zero. The

619 regulator specified that the survey be able to detect an average excess of even 1,500 cpm being

620 released. This value is the DL. The UBGR is set equal to the DL (i.e., 1,500 cpm), with $\beta =$

621 0.025. The owner of the ductwork felt that there was very little if any radioactivity above

622   background present, and was willing to set $\alpha = 0.20$. The estimated standard deviation of the

623   measurements, σ, was 375 cpm. The relative shift is $\Delta/\sigma = (UBGR - LBGR)/\sigma =$

624   $(1,500 - 0)/375 = 4$.

625   The sample size needed for the WRS test can be found in Table A.2b. The result is nine

626   measurements in each survey unit and nine in each reference material $\alpha/2 = 0.10$, and $\beta = 0.025$,

627   and $\Delta/\sigma = 4$. The data were obtained as in Section 6.6.3. Table 6.4 lists the gross count rate data

628   obtained. These data were reanalyzed using Scenario B and the results are shown in Table 6.5.

629   **Table 6.5 Scenario B WRS Test Example Data**

Data (cpm)	Area	Adjusted Data	Ranks	Survey Unit Ranks
2180	R	2180	11	0
2398	R	2398	13	0
2779	R	2779	16	0
1427	R	1427	1	0
2738	R	2738	15	0
2024	R	2024	6	0
1561	R	1561	2	0
1991	R	1991	5	0
2073	R	2073	8	0
2039	S	2039	7	7
3061	S	3061	17	17
3243	S	3243	18	18
2456	S	2456	14	14
2115	S	2115	9	9
1874	S	1874	4	4
1703	S	1703	3	3
2388	S	2388	12	12
2159	S	2159	10	10
		Sum =	171	94

630   In the "Area" column, the code "R" denotes a reference material measurement and "S" denotes a

631   survey unit measurement. The adjusted data would be obtained by subtracting the LBGR from

632   the survey unit measurements (see Section 6.6.2, Step 1), but since the LBGR is zero, no

633   adjustment is needed. The ranks of the adjusted data range from 1 to 18, since there are a total of

634   9+9 measurements (see Section 6.6.2, Step 2). Note that the sum of all of the ranks is still

635   18(18+1)/2 = 171. Checking this value with the formula in Step 5 of Section 6.6.2 is

636   recommended to guard against errors in the rankings.

637   The total of the ranks belonging to the survey unit measurements is 94. This is compared with

638   the entry for the critical value of 100 in Table A.4 for $\alpha = 0.10$, with $n = 9$ and $m = 9$. Since the

639   sum of the reference material ranks is less than the critical value, the null hypothesis (i.e., that

640   there is no detectable radioactivity above background) is not rejected, and the ductwork may be

641   released if the Quantile test is passed.

642   ## 6.7   Quantile Test

643   The Quantile test was developed to detect differences between the surveyed M&E and the

644   reference material that consist of a shift to higher values in only a fraction of the surveyed M&E.

645   The Quantile test is only performed when Scenario B is used, and only if the null hypothesis is

646   not rejected for the WRS test. Using the Quantile test, in tandem with the WRS test, results in

647   higher power to identify M&E that do not meet the disposition criterion than either test by itself.

648   Apply the Quantile test as follows:

649   1. Calculate α_Q ($\alpha_Q = \alpha/2$).

650   2. Obtain the adjusted survey unit measurements, Z_i, by subtracting the LBGR from each

651   survey unit measurement, Y_i. $Z_i = Y_i$ - LBGR.

652   3. The n adjusted survey unit measurements, Z_i, and the m reference material measurements,

653   X_i, are pooled and ranked in order of increasing size from 1 to N, where $N = m+n$.

654   4. If several measurements are tied (i.e., have the same value), they are all assigned the
655   mean rank of that group of tied measurements.

656   5. Look up the values for r and q in Table A.5 based on the number of measurements in the
657   survey unit (n), the number of measurements in the reference area (m), and α_Q. The
658   operational decision described in the next step is made using the values for r and q.

659   6. If q or more of the r largest measurements in the combined ranked data set are from the
660   survey unit, the null hypothesis is rejected.

661   This form of the Quantile test gives only approximate results, since Table A.5 provides a limited
662   number of combinations of n, m, and α_Q. It is recommended that several combinations of n, m
663   and α_Q be considered when interpreting the results of the Quantile test. Sections 7.2 and 7.3 of
664   NUREG-1505 (NRC 1998a) provide additional guidance on interpreting the results of the
665   Quantile test.

666   As an example, the Quantile test can be applied to the Class 2 Metal Ductwork example of
667   section 6.6.4. Using $n = 9$, $m = 9$, and $\alpha_Q = 0.10$, the nearest entry in Table A.5d has for $r = 3$
668    $q = 3$ with $\alpha_Q = 0.105$ when $n = 10$ and $m = 10$. This means that all three of the highest
669   measurement would have to be from the survey unit in order to reject the null hypothesis. From
670   Table 6.5, one can see that the two largest measurements are from the survey unit, but the third
671   largest is from the reference area. Since the ductwork has passed both the WRS and the Quantile
672   test in the Scenario B example, one would conclude that it could be released from radiological
673   controls.

674   ## 6.8 Evaluate the Results: The Decision

675   Once the data and results of the tests are obtained, the specific steps required to make a
676   disposition decision depends on the procedures approved by the regulator. The following
677   considerations are suggested for the interpretation of the test results with respect to the
678   disposition criteria. Note that the tests need not be performed in any particular order.

679 **6.8.1   Interpret Data for Each Survey Type**

680   The interpretation of results from the data evaluation or statistical test is the decision to reject or

681   not to reject the null hypothesis. For some of the survey designs the decision is straightforward,

682   while for other designs the interpretation is more complex. Figures 6.3 and 6.4 summarize the

683   interpretation of results.

684   6.8.1.1   Compare Results to the UBGR

685   The process for interpreting results compared to the UBGR depends on the action level used to

686   develop the survey design.

687   If the action level is zero or background, Scenario B must be used:

688   •   Compare every measurement result to the critical value corresponding to the required
689   scan MDC.
690   •   If all results are below the critical value, the M&E demonstrate compliance with the
691   disposition criterion.
692   •   Any results that exceed the critical value provide evidence of radionuclide
693   concentrations or radioactivity levels exceeding the disposition criteria, so the M&E
694   do not demonstrate compliance with the release criterion.

695   If the action level is not zero or background:

696   •   Compare every measurement result to the critical value corresponding to the UBGR.
697   •   If all results are below the critical value, the M&E demonstrate compliance with the
698   disposition criterion.
699   •   Any results that exceed the critical value provide evidence of radionuclide
700   concentrations or radioactivity levels exceeding the disposition criteria, so the M&E
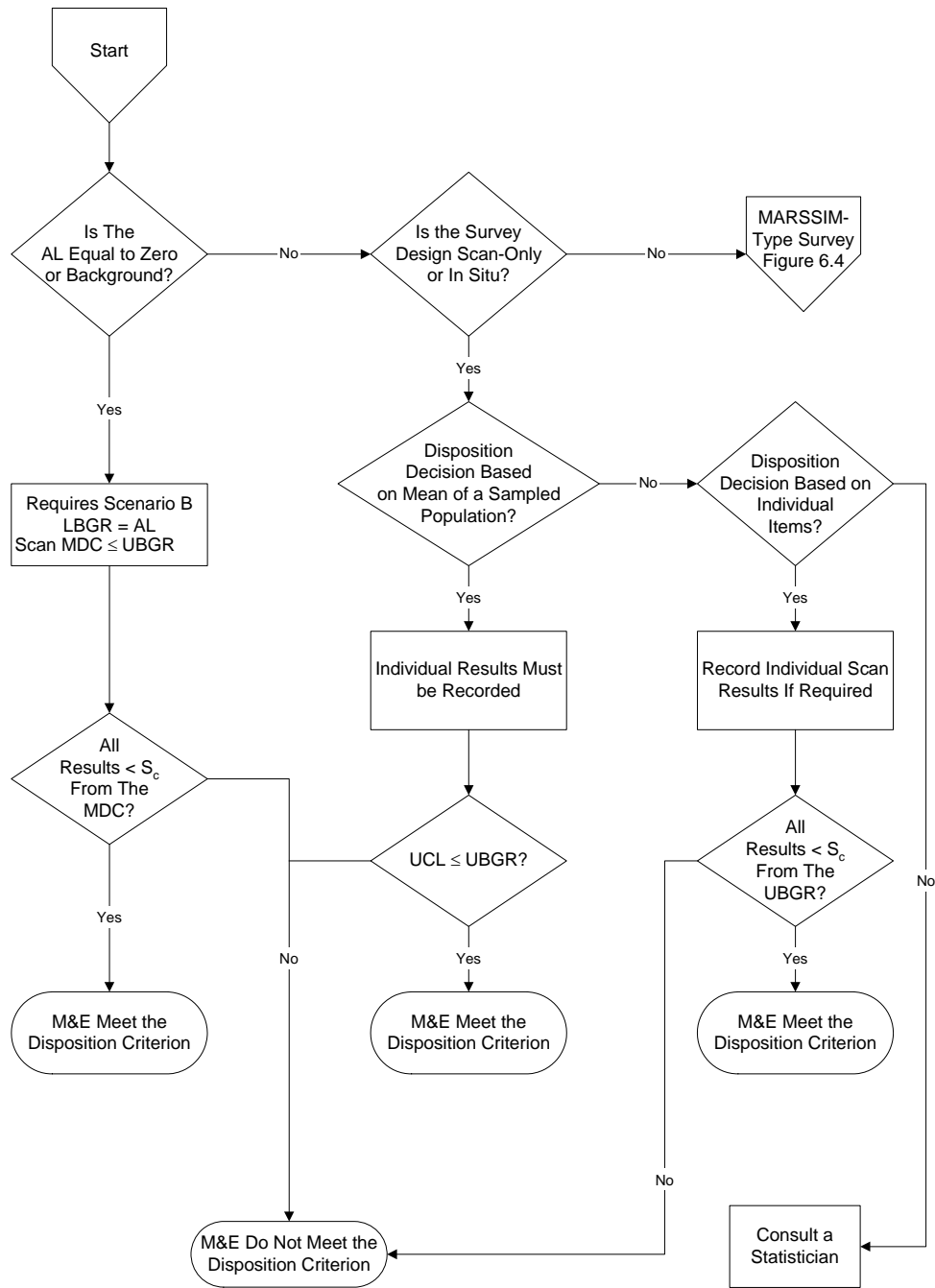701   do not demonstrate compliance with the release criterion.

702

703   **Figure 6.3 Interpretation of Survey Results for Scan-Only and In Situ Surveys**
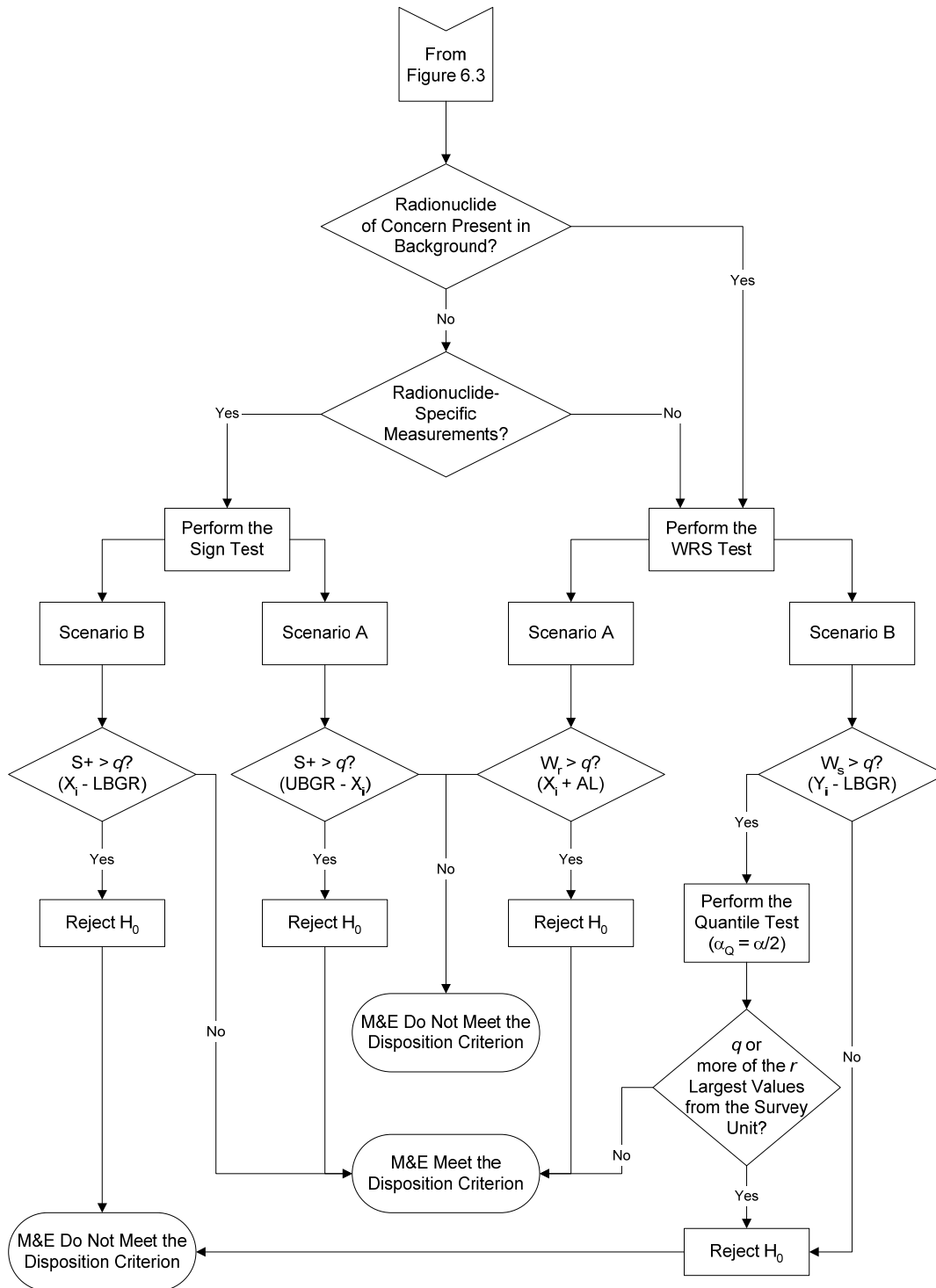
704
705 **Figure 6.4 Interpretation of Results for MARSSIM-Type Surveys**

706   Scan-only results are usually available as the data are collected. This real-time availability of

707   results allows the surveyor to make decisions as the data are collected. M&E that exceed the

708   action level can be identified and segregated during implementation of the survey. This "clean

709   as you go" approach to surveys is only applicable for Class 1 surveys (i.e., 100% of M&E are

710   measured) where there is high confidence in the quality and accuracy of detection decisions

711   around the UBGR. Extensive documentation of the measurement process, previous applications

712   of the process to the same or similar M&E, and verification of MDCs and MQCs is generally

713   necessary to implement a "clean as you go" survey design.

714   6.8.1.2   Compare Results Using an Upper Confidence Limit

715   When decisions are made based on the mean of a sampled population, the survey results should

716   be evaluated by comparison to an upper confidence limit:

717   • Compare every measurement result to the critical value corresponding to the UBGR.

718   • If all results are below the critical value, the M&E demonstrate compliance with the
719   disposition criterion.

720   • If any results are above the critical value, calculate the upper confidence limit (Section
721   6.4.1).

722   • If the upper confidence limit is less than the UBGR, the M&E demonstrate
723   compliance with the disposition criterion.

724   • If the upper confidence limit exceeds the UBGR, the M&E do not demonstrate
725   compliance with the disposition criterion.

726   • Investigate measurements exceeding the UBGR.

727   • Results above the UBGR trigger a reevaluation of classification as Class 2.

728   • Results above the MDC trigger a reevaluation of classification as Class 3.

729 6.8.1.3 MARSSIM-Type Surveys

730 The process for evaluating MARSSIM-type survey results is more complicated. This process is

731 explained in more detail in MARSSIM Section 8.5.

732   • Calculate the test statistics (see Section 6.5.1, 6.6.1, 6.6.2, and 6.7).

733   • Look up the critical value in the appropriate statistical table in Appendix A.

734   • Evaluate the results of the statistical test as described in Figures 6.3 and 6.4.

735   • Evaluate individual results using the elevated measurement comparison (EMC).

736   • M&E must pass the statistical test and the EMC (if applicable) to demonstrate

737   compliance.

738 If the null hypothesis is rejected under Scenario A, there is sufficient evidence to show the

739 median radionuclide concentrations or radiation levels are below the disposition criterion. Under

740 Scenario B, failing to reject the null hypothesis means there is insufficient evidence to overturn

741 the initial assumption the M&E demonstrate compliance with the disposition criterion.

742 If the null hypothesis is rejected under Scenario B, additional investigations are required to

743 determine the final disposition of the M&E (see Section 6.8.2). Failure to reject the null

744 hypothesis under Scenario A also requires additional investigations.

745 **6.8.2   Investigate Causes for Survey Unit Failures**

746 When M&E fail to demonstrate compliance with the disposition criterion, the first step is to

747 review and confirm the data that led to the decision. Once this is done, the DQO Process can be

748 used to evaluate potential problem areas leading to failure. If the level of radioactivity on or in

749 some Class 1 M&E exceeds the UBGR, the simplest solution might be to segregate those items

750 for a different disposition decision (see Section 6.8.1.1 on "clean as you go" surveys).

751 Sometimes activity in excess of background can be removed from the M&E followed by re-

752 evaluation or re-survey. In other cases, a less restrictive disposition option (e.g., disposal as

753 radioactive waste) may be selected. If such a situation were encountered in evaluating Class 2 or

754   Class 3 M&E, the classification would be questioned and the M&E would be reclassified and

755   surveyed as Class 1 M&E. This may also bring other classification decisions into question.

756   As a general rule, it may be useful to anticipate possible modes of failure. These can be

757   formulated as the problem to be solved using the DQO Process. Once the problem has been

758   stated, the decision concerning the failing survey unit can be developed into a decision rule. For

759   example, decide whether to attempt to remove the radioactivity or simply segregate certain types

760   of M&E for low-level waste disposal. Next, determine the additional data, if any, needed to

761   document that a survey unit where pieces with elevated measurements have been removed or

762   areas of added activity removed demonstrates compliance with the disposition criterion.

763   Alternatives to resolving the decision rule should be developed for each type of M&E that may

764   fail the surveys. These alternatives can be evaluated against the DQOs, and a disposition survey

765   design that meets the objectives of the project can be selected.

766   ## 6.9   Document the Disposition Survey Results

767   Documentation of survey results is an important part of the disposition survey process. The form

768   of this documentation can vary greatly depending on the survey objectives and regulatory or

769   administrative requirements. Documentation of disposition survey results should be considered

770   during survey design to ensure adequate records are provided during implementation. Generally,

771   survey documentation requirements are provided as part of the documented survey design.

772   Documented items may include:

773   • A description of the final disposition, such as disposal in a landfill, return to

774   manufacture for refurbishment, sold as salvage, recycled as ferrous metal, etc.

775   • A release statement to the transport carrier and recipient of the material indicating

776   that the M&E described in the bill of laden meet(s) applicable State and Federal

777   regulations.

778   • Results of QC measurements made during the conduct of release surveys and
779   confirmation of compliance with facility SOPs and action levels.

780   If the disposition survey is a routine survey documented in an SOP, the documentation
781   requirements for the actual measurement results may be minimal or non-existent. For example,
782   routine surveys performed to clear M&E from a facility may require documentation that the
783   instruments were calibrated and functioning properly and that trained personnel were on duty to
784   perform the surveys. Quality assurance reviews and audits would be performed periodically
785   (typically under a separate SOP) to document that the clearance surveys were being performed
786   properly and that no M&E were cleared without first being surveyed. These records would
787   document that properly trained personnel had adequately surveyed all M&E leaving the facility
788   using properly functioning instruments. Documentation of individual measurement results may
789   not be required or necessary.

790   If the survey is not routine, significantly more documentation may be required. This
791   documentation should provide a complete and unambiguous record of the radiological status of
792   the M&E relative to the selected action levels. In addition, sufficient data and information
793   should be provided to enable an independent evaluation of the survey results, including repeating
794   measurements at some future time. The documentation should comply with all applicable
795   regulatory requirements. Additional information on documentation is provided in Section 2.5,
796   Section 3.6, Section 4.5, MARSSIM Sections 3.8 and 8.6, and MARSSIM Chapter 5.